

Contents lists available at [ScienceDirect](http://ScienceDirect.com)

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on conditional variance and characterization of probability distributions

Piotr Jaworski^a, Marcin Pitera^{b,*}^a Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland^b Institute of Mathematics, Jagiellonian University, Łojasiewicza 6, 30-348, Cracow, Poland

ARTICLE INFO

Article history:

Received 25 January 2020

Received in revised form 16 April 2020

Accepted 16 April 2020

Available online 24 April 2020

MSC:

62E10

60E05

Keywords:

Conditional variance

Characterization

Censored variance

Trimmed variance

Integrated quantile function

Quantile division

ABSTRACT

In this note we prove a novel characterization result stating that any distribution is determined uniquely up to an additive constant by its conditional variance function where the conditioning is based on double quantile trimming. We also outline potential statistical applications of the proposed characterization.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Characterization results for probability distributions are important part of statistics and probability applications. This includes generic distribution classifiers like characteristic or mean residual life functions as well as specific identification methods like independence of mean and variance estimators for Gaussian distributions; see [Galambos and Kotz \(2006\)](#) and [Ahsanullah \(2017\)](#), and references therein. In this note, we focus on the former and show that the information about doubly quantile censored variance function is sufficient to uniquely determine the distribution up to an additive constant.

It should be noted that there is a reach literature linked to characterizations based on various types on conditional first moments; see e.g. [Ruiz and Navarro \(1996\)](#), [Khan \(2010\)](#), and [Ahsanullah et al. \(2016\)](#). In particular, in [Navarro et al. \(1998\)](#) it is shown that the doubly censored mean function given by $m(x, y) = \mathbb{E}[X \mid x < X < y]$ could be used to uniquely determine the distribution. What is more, instead of pre-defined values, one could consider the integrated quantile functions for characterization; see Theorem 3 in [Khan \(2010\)](#) for details.

The characterization theorems based on conditional second (and higher) moments have been studied in the literature only in specific contexts. For example, in [Unnikrishnan Nair and Sudheesh \(2010\)](#) the authors study how the properties of truncated variance function could result in characterizations for specific classes of non-negative absolutely continuous random variables satisfying certain properties; see [Unnikrishnan Nair and Sudheesh \(2006\)](#) where the required condition, given in Theorem 2.1(iv), is discussed in details. Also, in [El-Arshy \(2005\)](#) the conditional variance characterization in a

* Corresponding author.

E-mail address: marcin.pitera@uj.edu.pl (M. Pitera).

specific context of some discrete probability distributions is given. Finally, it should be noted that the potential usage of truncated moments as classifiers has been communicated in the literature (e.g. in [Laurent, 1974](#)) but we found no direct treatment of this property and the discussion about its potential application.

While the conditional variance function with quantile set trimming seems to be a natural (local) extension of standard variance, it is not considered in the literature as a benchmark framework. This is quite surprising, as the conditional second moments seem to be more natural (e.g. for engineering applications) compared to higher-order moment analysis, e.g. when the tail structure is assessed. In fact, it was shown recently in [Jelito and Pitera \(2018\)](#) that a simple test based on conditional second moments outperforms most of the popular benchmark methods when normality testing is considered. More explicitly, the statistical test power for various choices of popular alternatives (t-student, logistic, and Cauchy distributions) was shown to be bigger compared to reference normality tests based on Jarque–Bera, Anderson–Darling, or Shapiro–Wilk statistics; see [Jelito and Pitera \(2018, Table 3\)](#) for more details. See also [Hebda-Sobkowicz et al. \(2020\)](#) where a similar approach has been used for the local damage detection in mining (ore deaggregation) process.

The characterization result presented in this note shows that conditional variances might be used for efficient distribution identification and goodness-of-fit testing. In particular, it shows that one could develop efficient statistical testing framework, by controlling the number of included conditional sets with the sample size.

2. Preliminaries

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and let $L^0 := L^0(\Omega, \Sigma, \mathbb{P})$ denote the set of all (a.s. identified) random variables. For any $X \in L^0$ and $A \in \Sigma$, such that $\mathbb{P}[A] \neq 0$, we use

$$\text{Var}[X|A] := \mathbb{E}[(X - \mathbb{E}[X|A])^2|A] \quad (2.1)$$

to denote (possibly infinite) conditional variance of X on A ; all regularity conditions are taken for granted. For brevity, for any $X \in L^0$ and $0 \leq a < b \leq 1$ we define a quantile conditioned variance

$$V_X(a, b) := \text{Var}[X|A_X(a, b)], \quad (2.2)$$

where $A_X(a, b) := \{\omega \in \Omega : X(\omega) \in [Q_X(a), Q_X(b)]\}$ is the quantile set with the lower quantile function $Q_X : [0, 1] \rightarrow [-\infty, +\infty]$ given by

$$Q_X(u) := \inf\{x : u \leq F_X(x)\}, \quad u \in [0, 1], \quad Q_X(1) := \sup\{x : F_X(x) < 1\}. \quad (2.3)$$

It should be noted that $V_X(a, b)$ is well defined as for $0 \leq a < b \leq 1$, we get

$$\mathbb{P}[A_X(a, b)] = F_X(Q_X(b)) - F_X^l(Q_X(a)) \geq b - a > 0, \quad (2.4)$$

where $F_X^l(t) := \mathbb{P}(X < t)$, $t \in \mathbb{R}$, denotes the Kolmogorov distribution function. Indeed, recalling that Q_X is the left-continuous generalized inverse of the cumulative distribution function, and for any $u \in [0, 1]$ we have $F_X^l(Q_X(u)) \leq u \leq F_X(Q_X(u))$, we get (2.4). Also, it is worth noting that for $u \in [0, 1]$ we get

$$F_X(Q_X(u)) - F_X^l(Q_X(u)) = \mathbb{P}[X = Q_X(u)]. \quad (2.5)$$

Finally, note that the quantile conditional variance function given in (2.2) is defined up to an additive constant, i.e. for any fixed $X \in L^0$, $0 \leq a < b \leq 1$, and $c \in \mathbb{R}$, we get $A_X(a, b) = A_{X+c}(a, b)$, and $V_X(a, b) < \infty$ if additionally $0 < a$ and $b < 1$.

3. Main result

In this section we state and prove the main result of this note, i.e. that the information about quantile-based conditional variance is sufficient to characterize the distribution of X up to an additive constant.

Theorem 3.1. *Let X, Y be any random variables such that for $0 \leq a < b \leq 1$ we have $V_X(a, b) = V_Y(a, b)$. Then, there exists $c \in \mathbb{R}$ such that $F_X(t) = F_{Y+c}(t)$, $t \in \mathbb{R}$, i.e. the laws of X and Y coincide almost surely up to an additive-constant.*

Proof. Following the second proof of Theorem 14.1 in [Billingsley \(2008\)](#) we observe that a random variable X has the same distribution as the random variable $Q_X(U)$, where $U(\omega) := \omega$ is a uniformly distributed random variable defined on the standard probability space $((0, 1), \mathcal{B}((0, 1)), \lambda)$ and λ is the Lebesgue measure.

For any $0 \leq a < b \leq 1$, let $\tilde{A}(a, b) := \{\omega \in [0, 1] : \omega \in [F_X^l(Q_X(a)), F_X(Q_X(b))]\}$. Recalling (2.4) and noting that $\tilde{A}(a, b) = \{\omega \in [0, 1] : Q_X(\omega) \in [Q_X(a), Q_X(b)]\} \cup \{F_X^l(Q_X(a))\}$, we get

$$\begin{aligned} \mathbb{E}[X|A_X(a, b)] &= \mathbb{E}[Q_X(U)|\tilde{A}(a, b)] = \frac{1}{\mathbb{P}[A_X(a, b)]} \int_{F_X^l(Q_X(a))}^{F_X(Q_X(b))} Q_X(u) du, \\ \mathbb{E}[X^2|A_X(a, b)] &= \mathbb{E}[Q_X(U)^2|\tilde{A}(a, b)] = \frac{1}{\mathbb{P}[A_X(a, b)]} \int_{F_X^l(Q_X(a))}^{F_X(Q_X(b))} Q_X(u)^2 du. \end{aligned}$$

Consequently, using the property $V_X(a, b) = \mathbb{E}[X^2|A_X(a, b)] - \mathbb{E}^2[X|A_X(a, b)]$, we get

$$V_X(a, b) = \frac{1}{2\mathbb{P}[A_X(a, b)]^2} K_X(F_X^l(Q_X(a)), F_X(Q_X(b))), \quad (3.1)$$

where $K_X : [0, 1]^2 \rightarrow [-\infty, +\infty]$ is an exchangeable function ($K(x, y) = K(y, x)$) given for $x \leq y$ by

$$\begin{aligned} K_X(x, y) &:= 2(y-x) \int_x^y Q_X^2(u) du - 2 \left(\int_x^y Q_X(u) du \right)^2 \\ &= \int_x^y \int_x^y (Q_X^2(u) + Q_X^2(v)) dudv - \int_x^y \int_x^y 2Q_X(u)Q_X(v) dudv \\ &= \int_x^y \int_x^y (Q_X(u) - Q_X(v))^2 dudv. \end{aligned} \quad (3.2)$$

Now, let ΔK_X denote the mixed second difference of K_X ,

$$\Delta K_X(x_1, x_2, y_1, y_2) := K_X(x_1, y_1) + K_X(x_2, y_2) - K_X(x_1, y_2) - K_X(x_2, y_1). \quad (3.3)$$

For $0 \leq x_1 < x_2 < y_2 < y_1 \leq 1$ we obtain

$$\Delta K_X(x_1, x_2, y_1, y_2) = 2 \int_{x_1}^{x_2} \int_{y_2}^{y_1} (Q_X(u) - Q_X(v))^2 dudv. \quad (3.4)$$

Now, we show that the intervals where the quantile function Q is constant are determined by the conditional variance V . Let $\varphi, \psi : [0, 1] \rightarrow [0, 1]$ be given by $\varphi(0) = 0$, $\psi(1) = 1$ and

$$\varphi(u) := \sup\{x : 0 \leq x < u, V_X(x, u) > 0\}, \quad \psi(u) := \inf\{x : 1 \geq x > u, V_X(u, x) > 0\}.$$

For $u \in [0, 1]$ such that $\varphi(u) < \psi(u)$, the quantile function Q_X is constant on $(\varphi(u), \psi(u)]$, and for any $v \in (0, 1)$ such that $v \neq u$ and $Q_X(v) = Q_X(u)$, we get $v \in [\varphi(u), \psi(u)]$. Therefore, for $u \in (0, 1)$, we get $\varphi(u) = F_X^l(Q_X(u))$ and $\psi(u) = F_X(Q_X(u))$. Consequently, recalling (3.1), for $0 < a < b < 1$, we get

$$K_X(\varphi(a), \psi(b)) = (\psi(b) - \varphi(a))^2 V_X(a, b). \quad (3.5)$$

Furthermore, for $\varphi(u) > 0$ and small $\varepsilon > 0$, we get $Q_X(\varphi(u) - \varepsilon) < Q_X(u)$, which implies

$$\varphi(u) = F^l(Q_X(u)) \geq F(Q_X(\varphi(u) - \varepsilon)) \geq \varphi(u) - \varepsilon. \quad (3.6)$$

Since $F(Q_X(\varphi(u) - \varepsilon)) = \psi(\varphi(u) - \varepsilon)$, we get $\lim_{v \rightarrow \varphi(u)^-} \psi(v) = \varphi(u)$. Next, since $K_X(a, b)$ is continuous, for b such that $\varphi(b) > 0$, we get $K_X(a, \varphi(b)) = \lim_{v \rightarrow \varphi(b)^-} K_X(a, \psi(v))$ and, due to (3.5), for $\varphi(a) < \varphi(b)$, we obtain

$$K_X(\varphi(a), \varphi(b)) = \lim_{v \rightarrow \varphi(b)^-} (\psi(v) - \varphi(a))^2 V_X(a, v). \quad (3.7)$$

Similar reasoning is true for any u such that $\psi(u) < 1$. Noting that for small $\varepsilon > 0$ we have $Q_X(u) < Q_X(\psi(u) + \varepsilon)$, we get $\lim_{v \rightarrow \psi(u)^+} \varphi(v) = \psi(u)$. Thus, for a such that $\psi(a) < \psi(b)$, we get

$$K_X(\psi(a), \psi(b)) = \lim_{v \rightarrow \psi(a)^+} (\psi(b) - \varphi(v))^2 V_X(v, b). \quad (3.8)$$

Note that if $\psi(a) = \psi(b)$, then $K(\psi(a), \psi(b)) = 0$, and if $\varphi(b) = \varphi(a)$, then $K(\varphi(a), \varphi(b)) = 0$.

Now, if $\varphi(u) < \psi(u)$, then the quantile function $Q_X(\cdot)$ is constant on $(\varphi(u), \psi(u))$, so both $K_X(\cdot, \psi(u))$ and $K_X(\varphi(u), \cdot)$ are linear on $(\varphi(u), \psi(u))$. Combining all facts, we get that integrals K_X and ΔK_X are determined by the conditional variance V_X .

Now, let us assume that Y is such that $V_X(a, b) = V_Y(a, b)$ for all $0 \leq a < b \leq 1$. Since Q_X and Q_Y are left-continuous and nondecreasing, the equality $\Delta K_X = \Delta K_Y$ implies that $Q_X(b) - Q_X(a) = Q_Y(b) - Q_Y(a)$ for all $a, b \in [0, 1]$. This concludes the proof. \square

Theorem 3.1 could be easily extended to the multivariate case e.g. by using information about conditional variances for all linear combination of marginal random variables. In the following theorem we use $\langle \cdot, \cdot \rangle$ to denote the standard Euclidean inner product operator.

Theorem 3.2. Let X, Y be any n -dimensional random vectors such that for $0 \leq a < b \leq 1$ and $\alpha \in \mathbb{R}^n$ we have $V_{\langle \alpha, X \rangle}(a, b) = V_{\langle \alpha, Y \rangle}(a, b)$. Then, there exists $c \in \mathbb{R}^n$ such that $F_X(t) = F_{Y+c}(t)$, $t \in \mathbb{R}^n$, i.e. the laws of X and Y coincide almost surely up to an additive shift.

The proof of **Theorem 3.2** follows directly from **Theorem 3.1** combined with Theorem 19 from Galambos (1995). To conclude, let us present two simple remarks which outline potential application of **Theorem 3.1**; similar remarks are true for the multivariate case.

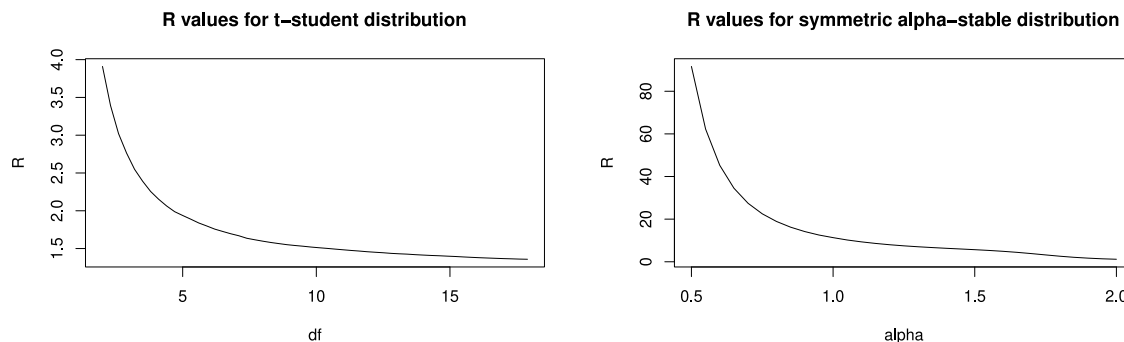


Fig. 1. The plot illustrates the values of $R = V_X(0.1, 0.3)/V_X(0.3, 0.7)$ under the assumption that X has t-student (left) or symmetric α -stable (right) distribution. R is presented as a function of the underlying parameters: df for t-student (left) and α for symmetric α -stable (right). The values were obtained using Monte Carlo samples of size 10 000 000. In both cases R is a decreasing function of the underlying parameter.

Remark 3.3 (*Statistical Goodness-of-fit Testing*). As quantile-based conditional variances are easy to estimate and could be used to uniquely classify the distribution (up to an additive constant), they are a natural candidate for goodness-of-fit (shape) statistical testing. In practical applications, it is reasonable to choose a fixed set of specific quantile conditioned sets and then compare conditional variances with the theoretical variances coming from the reference distribution. By introducing various quantile splits and appropriate ratios one might check certain distributional properties rather than the full fit. For example, the comparison of $\hat{V}_X(a, b)$ and $\hat{V}_X(1 - a, 1 - b)$ for any $0 \leq a < b \leq 1$ might be used to test distribution symmetry. Also, for $a < 0.5$, the tail set conditional variances $\hat{V}_X(0, a)$ and $\hat{V}_X(1 - a, 1)$ might be compared with the central set conditional variance $\hat{V}_X(a, 1 - a)$ in order to assess heaviness of the distribution tail.

In fact, exemplary normality testing framework based on conditional variance estimation has been recently introduced in Jelito and Pitera (2018). Using the fact that $V_X(0, 0.2) = V_X(0.2, 0.8) = V_X(0.8, 1)$ for Gaussian random variables we can define the test statistic

$$N := \frac{\sqrt{n}}{1.8} \cdot \frac{\hat{V}_X(0, 0.2) + \hat{V}_X(0.8, 1) - 2\hat{V}_X(0.2, 0.8)}{\hat{V}_X(0, 1)}, \quad (3.9)$$

where $\hat{V}_X(a, b)$ refers to sample conditional variance constructed by sorting the sample, taking appropriate subset of observation, and applying standard sample variance estimator.¹ In Jelito and Pitera (2018), it is shown that the power of related normality test for various choices of popular symmetric alternatives (e.g. t-student, logistic, and Cauchy distributions) is surprisingly big. In particular, test statistic N outperforms popular alternatives like Jarque–Bera, Anderson–Darling, or Shapiro–Wilk tests for samples of size 20, 50, 100, and 250; see Jelito and Pitera, 2018, Table 3 for details. Also, it is easy to show that N is asymptotically normal.

Remark 3.4 (*Parameter Fitting*). Conditional variances could be also used for parameter fitting. While being relatively simple to establish, the framework based on conditional second moments is much more flexible compared e.g. to method of moments. This is due to the fact that one could consider multiple choices of quantile intervals (a, b) and take their linear combinations; note that sample quantile conditional variance estimators are consistent. To illustrate this, let us consider the ratio $R := V_X(0.1, 0.3)/V_X(0.3, 0.7)$ for two distribution families: t-student and symmetric α -stable; see Ahsanullah (2017) for details. In Fig. 1, we present the values of R as a function of degrees of freedom (df) and stability index (α) parameters, respectively; note that R is invariant to affine transformations of X . One could see that in both cases R is monotone wrt. parameter change, so that is could be used for parameter identification.

References

- Ahsanullah, M., 2017. Characterizations of Univariate Continuous Distributions. Atlantis Press.
- Ahsanullah, M., Shakil, M., Kibria, B., 2016. Characterizations of continuous distributions by truncated moment. J. Modern Appl. Statist. Methods 15 (1), 17.
- Billingsley, P., 2008. Probability and Measure. John Wiley & Sons.
- El-Arshy, S., 2005. A conditional variance characterization of some discrete probability distributions. Statist. Papers 46 (1), 31–45.
- Galambos, J., 1995. Advanced Probability Theory, Vol. 10. CRC Press.
- Galambos, J., Kotz, S., 2006. Characterizations of Probability Distributions: A Unified Approach with an Emphasis on Exponential and Related Models, Vol. 675. Springer.

¹ This (approximate) ratio for which the conditional variance equality is true might be linked to so called 20/60/20 Rule; we refer to Jaworski and Pitera (2016, 2017) for discussion.

- Hebda-Sobkowicz, J., Zimroz, R., Wylomańska, A., 2020. Selection of the informative frequency band in a bearing fault diagnosis in the presence of non-Gaussian noise – comparison of recently developed methods. *Appl. Sci.* 10 (8), 2657.
- Jaworski, P., Pitera, M., 2016. The 20-60-20 rule. *Discrete Contin. Dyn. Syst. Ser. B* 21 (4).
- Jaworski, P., Pitera, M., 2017. A note on conditional covariance matrices for elliptical distributions. *Statist. Probab. Lett.* 129, 230–235.
- Jelito, D., Pitera, M., 2018. New Fat-Tail Normality Test Based on Conditional Second Moments with Applications to Finance. Available at SSRN 3303132.
- Khan, M.J.S., 2010. Moments and Characterization of Probability Distributions (Ph.D. thesis). Aligarh Muslim University.
- Laurent, A.G., 1974. On characterization of some distributions by truncation properties. *J. Amer. Statist. Assoc.* 69 (347), 823–827.
- Navarro, J.M., Ruiz, J., Zoroa, N., 1998. A unified approach to characterization problems using conditional expectations. *J. Statist. Plann. Inference* 69 (2), 193–207.
- Ruiz, J.M., Navarro, J., 1996. Characterizations based on conditional expectations of the doubled truncated distribution. *Ann. Inst. Statist. Math.* 48 (3), 563–572.
- Unnikrishnan Nair, N., Sudheesh, K.K., 2006. Characterization of continuous distributions by variance bound and its implications to reliability modeling and catastrophe theory. *Comm. Statist. Theory Methods* 35 (7), 1189–1199.
- Unnikrishnan Nair, N., Sudheesh, K.K., 2010. Characterization of continuous distributions by properties of conditional variance. *Stat. Methodol.* 7 (1), 30–40.